

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

Evaluation of PATB and Its Use for  
Selecting Personnel

Summary of Major Findings and Recommendations

p.1 [ Our assigned tasks were to review the evidence for validity and reliability of PATB and its use in the Agency for selecting personnel for professional positions, to assess its fairness for use with all applicants for professional positions, and to make recommendations for improvement in the tests and their use. Our assignment did not p.2 [ include evaluation of the intensive psychological assessment procedures or any other procedures used to select personnel for jobs in the Agency, and these also require validation.

Major Findings

A. Use of PATB to Select Personnel

1. If scores on PATB are used at all to help select personnel, they enter the decision-making process only after a unit has indicated an interest in an applicant listed on the Applicant File Listing circulated by the Skills Bank. If no one indicates an interest in an applicant on this list within

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

Skills Bank. If no one indicates an interest in an applicant on this list within 10 days, then no report of that applicant's scores on PATB is made.

2. Only about one-fourth of the supervisors in the units indicate that they give significant weight to the applicant's performance on PATB in making employment decisions.
3. About two thirds of professional employees who entered on duty between 1 October 1977 and August 1979 reported that they had taken PATB and one third reported that they had not.
4. There appears to be no consistent policy among the units on requiring applicants to take PATB.

B. Validity of PATB

1. There is no explicit rationale for PATB and no evidence that the initial construction of the test battery was based on a systematic analysis of professional jobs in the Agency. .
2. No studies have been done to establish the construct validity of the tests and scales that comprise PATB.

?  
taken from  
p. 17 of  
[redacted]

There are no validity data of any kind for the

odd patchwork  
of sentences  
that hang together  
in a schizophrenic  
way.

STATINTL

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

ADMINISTRATIVE-INTERNAL USE ONLY

BS

writing sample or for the Strong-Campbell Interest Inventory. Although there have been some attempts to validate certain items on the Biographical Information Inventory, evidence for its validity is lacking. *Chose to ignore it*

*We do not differentially validate some items while looking at an inventory. We either validate it (the entire inventory) or we do not.*

4. Four studies have been done on the validity of PATB for predicting success in foreign language training. The results of these studies indicate very little or no value of PATB for this purpose.

*These studies do not indicate this to me!*

- ✓ 5. Out of 16 studies of the job related validity (criterion-related validity), only 10 met at least minimal standards for reporting data and could be used to judge the validity evidence.

The evidence for job-related validity of PATB is *probably true* *opinion* *opinion* fragmentary, very weak and unconvincing. Correla-

tions of test scores with supervisory ratings of

job performance across similar jobs have been *opinion - not supported by the statistical evidence* inconsistent and generally low. Correlations of

test scores with similar or identical criteria of

job performance across jobs have also been *SAME* inconsistent and generally low.

*i.e. we chose only to consider the 10 we wanted to consider. Besides, if we reduced the # of studies claimed by PSS to meet rigid criteria, it makes us (the IG) look like we are the ones with the highest standards. 6.*

ADMINISTRATIVE-INTERNAL USE ONLY

ADMINISTRATIVE-INTERNAL USE ONLY

- ✓ 7. Equations generated by multiple regression analysis or discriminant analysis to predict job performance *we don't generate data here, we report it.* have been based on inadequate samples of small size and have not been cross-validated. *incorrect. redundant*

8. The samples used for studying job-related validity and to generate equations to predict job performance *Not true at all. IS has no way of knowing this, nor any way of substantiating such a claim.* have been composed solely or predominantly of white

*STATINTL*  
*21*  
*11* *reps*  
[REDACTED] *21* *reps* *They have been represented in exactly the proportions they represent in the real employee population.*  
males. Females have been inadequately represented in these samples. There is no evidence that minorities have been included in these samples.

9. The level of confidence with which the results of these studies have been reported and used to predict job performance of applicants greatly exceeds the level of confidence of the statistical data on validity and reliability. *Conversely, nor any evidence that they have not, as is the claim above.*

*I do not know what this means. If I read this as a statistician, I am immediately impressed with the significance of the reported validities. They must be very impressive indeed.*

- ✓ 10. The evidence on validity presented in the 10 studies does not meet minimum standards set by the American Psychological Association or the EEOC Uniform Guidelines on Employee Selection Procedures (1978).

*Specifically, where do they fall short? and how many of these were pre-1978 studies?*

ADMINISTRATIVE-INTERNAL USE ONLY

ADMINISTRATIVE-INTERNAL USE ONLY

C. Reliability of PATB

1. No reliability data for any of the tests and scales of PATB are available for minorities. *being done*
2. No reliability data for the work attitude scales are available for females.
3. The reliability of the writing sample or of the scoring of the writing sample has not been determined.
4. For white males and females, only 5 tests--Reading Vocabulary, Reading Comprehension, Arithmetic Problems, Numerical Operations and Interpretation of Data--have reliabilities that are at least minimally acceptable for making decisions about individuals. For all of the other tests and scales, the reliabilities are too low; thus the scores from them are too unstable and inaccurate for use in making decisions about individuals.

D. The Narrative Report

1. No written guidelines are used by the psychologists in writing the narrative report and the reports tend to vary considerably in unpredictable ways. The variations appear to be due as

*see p. 14 of appendix 2*

STATINTL

*they ( [redacted] ) 5 are talking about the report of the writing sample, not of the entire report.*

ADMINISTRATIVE-INTERNAL USE ONLY

ADMINISTRATIVE-INTERNAL USE ONLY

much or more to the personal idiosyncrasies of the psychologists as to differences in performance on PATB among applicants.

B.S.

2. The strong, confident recommendations to hire or not hire applicants for specific jobs or in specific units are not supported by the available evidence on validity of PATB.

wrong!

3. The scores from the Strong-Campbell Interest Inventory are being consistently misinterpreted in the narrative reports.

categorically wrong!

4. The reports of writing ability use ambiguous descriptors, are noncomparable from one report to another, and are inconsistent in the aspects of writing ability reported on.

E. Evidence of Unfairness

1. There is no evidence that studies of adverse impact as defined by EEOC have been done for PATB or for any other selection procedure used in the Agency.

2. Although we could find no direct evidence on bias or unfairness, the following factors indicate that there is serious potential for misuse or unfair use of PATB.

they sound disappointed

STATINTL

STATINTL

Verbatim from page 31 of report, report added that the word "serious" and the word "misuse"

ADMINISTRATIVE-INTERNAL USE ONLY

ADMINISTRATIVE-INTERNAL USE ONLY

- a. There is no evidence that minorities were represented in the groups used to norm the tests.
- b. There is no evidence that minorities have been included in the samples used to determine job-related validity. There is evidence that females have not been represented at all in some samples and are underrepresented in other samples.
- c. The equations being used to make recommendations for hiring employees have not been cross-validated. The sample used to generate these equations appears not to have included any minorities and to have had very few females.

or that they weren't

taken point by point from

STATINTL

31 d  
[redacted]  
repe

STATINTL

F. Other Findings

→ [redacted] calls this "Additional Comments," but adds that "we think the agency needs a good battery of selection tests...."

1. Units in the Agency are using other tests constructed within the units to select personnel. There is no evidence that these tests have been validated.
2. The tests and scales of PATB are over 20 years old. The content of some of the tests is dated and needs revision.

7  
ADMINISTRATIVE-INTERNAL USE ONLY

ADMINISTRATIVE-INTERNAL USE ONLY

Recommendations

We recommend that:

1. A systematic analysis of professional jobs in the Agency be done to determine the abilities needed to perform each job. On the basis of this analysis, it should be determined whether there is enough commonality in job requirements and demands to justify a single selection battery for use with applicants for professional jobs. If the answer is yes, and we think it will be, then a new battery should be constructed or the present PATB should be revised to appraise the abilities brought out by the job analyses.

STATINTL

from pp  
36-37 of  
[REDACTED]

2. Policy guidelines be set as to which applicants for professional jobs should be required to take PATB.

not established  
by PSS.  
who has  
responsibility?  
does not  
apply to PSS

3. The use of tests constructed in units but which have not been validated be controlled.

4. Equations to predict job performance not be used for minorities and women until both of these groups have been adequately represented in the samples used to generate the equations.

5. Equations to predict job performance for white males not be used until they have been cross-validated.

6. The use of the present narrative report prepared by psychologists be discontinued and instead a profile

ADMINISTRATIVE-INTERNAL USE ONLY



report or a completely standardized verbal description of performance be generated by computer.

7. The self-report instruments such as work attitude scales, temperament scales and biographical information not be used as a part of the basic battery to select employees.

They are subject to faking, call upon the individual to "testify against himself", lack demonstrated validity for performance of jobs, and have very low reliabilities.

8. The use of the Strong-Campbell Interest Inventory be discontinued because it lacks Agency norms and Agency validation, and is being consistently misinterpreted in the narrative reports.

9. Test results enter more actively into the selection process and the Applicant File Listing call attention to those applicants with superior scores on the cognitive tests.

10. Improved procedures for storage and retrieval of applicants' files be developed so that records of promising applicants, especially those with high abilities or special skills, can be considered for employment.

If no job analysis is done and the Agency decides to continue with the cognitive test portions of PATB, we

*We can tell if they're faking. Give us some credit for understanding an instrument we use everyday.*

*> How determined? These measures are expected to change over time!!*

*I agree! We have abilities scores as well as SCT data. Therefore, on abilities comments are quite appropriate.*

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

think the resultant "PATB" should be revised to improve its reliability and validity. We recommend that:

1. The Numerical Operations test be dropped. It is appraising speed of simple arithmetical computations which does not <sup>??</sup> seem to be related to Agency jobs.

*We are looking at style, at trade-offs between speed + accuracy, not just at arith skills.*

*This is important for ops work + for letting us print out who tends to do well or poorly in interviews.*

2. The Considerations test be dropped because it has very low reliability, needs to be hand-scored, and is not measuring an ability that appears relevant to Agency jobs.

3. The Interpretation of Data test be dropped because its content is dated. In its place, a new test be devised which would appraise the ability to synthesize information from separate and inconsistent sources.

4. The Vocabulary test be revised to exclude exotic words and to relate its content more closely to Agency jobs.

5. The Reading Comprehension test be revised replacing the literary reading selections with reading passages drawn from materials that must be read and comprehended by people performing Agency jobs.

6. The Arithmetic Problems test be revised replacing dated items.

7. The writing task be modified to make it correspond more closely to the type of writing done in professional jobs in the Agency.

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

8. The battery of tests be renormed on a representative sample of applicants making sure that minorities have the same proportional representation in the norm sample as they have in the applicant sample.

~~ADMINISTRATIVE-INTERNAL USE ONLY~~